Markov Models and Extensions for Land Cover Mapping in Aerial Imagery

Mohamed El Yazid Boudaren, Abdel Belaïd

Markov Models and Extensions for Land Cover Mapping in Aerial Imagery

Mohamed El Yazid Boudaren, IAENG member, Abdel Belaïd

***Abstract*— Markov models are well-established stochastic models for image analysis and processing since they allow one to take into account the contextual relationships between image pixels. In this paper, we attempt to methodically review the use of Markov models and their extensions for Land Cover mapping problem in aerial imagery according to available literature and previous research works. A new Markov model combining Markov random fields and hidden Markov models and inspired from the NSHP-HMM model, initially introduced for Handwritten Words Recognition is defined. New learning and labeling procedures are derived.**

***Index Terms*—Aerial Images, Land Cover Mapping, Markov models, Markov Random Fields MRF, Hidden Markov Models HMM.**

I. Introduction

Availability of accurate and up-to-date terrain cover information is crucial to many military and public applications. For instance, land cover maps are essential inputs for Combat simulators and agricultural models.

Due to their synoptic view and map like format, aerial images are viable sources for producing effective land cover classification.

Land cover mapping in aerial pictures can be seen as an image labeling problem since we need to assign to each pixel a missing label.

Image labeling is a topic of great importance for many computer vision systems. In general, three main approaches are used to solve computer vision problems: stochastic, structural and neural approaches.

Stochastic modeling is of a great importance to solve image labeling problem; especially when there is no direct deterministic link between labels and observable pixels (two pixels sharing the same characteristics may be assigned different labels) [12]. Consequently, there is evidence that stochastic modeling is the most suited to land cover mapping task given the large variations in aerial images [13].

Most stochastic systems incorporate Markov models which provide a basis for modeling contextual constraints in visual processing and interpretation.

The use of contextual information within image pixels means to go beyond the independence assumption between labels. In term of probability, the contextual constraints are expressed through local conditional probabilities.

In non-context situation (when independence assumption holds) the image observation joint probability is the product of local label observation probability.

Markov models were widely used in aerial image interpretation and segmentation: land cover mapping [1, 2, 8], remote sensing [9], Coast Line detection [7] and Forest Change detection [10].

In this paper, we aim to investigate which Markov modeling could yield best results for the task of segmenting high-resolution aerial images of rural areas into its constituent cartographic objects (fields, orchards, forests, lakes…).

The rest of the paper is organized as follows: section II defines the land cover mapping problem. Section III reviews the use of several Markov models for land cover mapping. In Section IV we remind the basics of the NSHP-HMM model we defined in [3]. In section V, we define a novel Markov model combining Dependency Tree Hidden Markov Model (DT-HMM) and NSHP-MRF. Conclusion remarks and future work are discussed in the last section.

II. Land Cover Mapping in Aerial Imagery

In this section, we formally define the so called land cover mapping problem.

Let us consider an aerial image S of size $T = m \times n$ pixels, where m and n represent the image length and width respectively. Each pixel $s \in S$ is described by an observable parameter $y_s \in V$. The symbol set $V = \{v_1 \dots v_M\}$ may correspond to a color space or any other characteristics.

The problem under consideration in this section is to allocate each of the image pixels $s \in S$ a missed (unobservable) label $x_s \in E$ where $E = \{e_1 \dots e_N\}$. All e_i are supposed beforehand known so that training can be performed on each of them taken alone. In this context, they correspond to natural object classes (fields, forests, lakes…).

The problem solution consists in deriving a class-map $x \in E^T$ from a given aerial image observation y.

We assume that aerial image S is of a resolution enough high that each pixel s belongs to only one natural object class.

The problem can be seen as a segmentation problem since neighboring pixels tend to belong to the same a region. The only difference is that besides segmenting the aerial image into distinguishable regions, we need to identify each region.

III. Markov Models for Land Cover mapping

A. Bayesian Naïve Segmentation

The problem resolution of the aerial image mapping described in section II can be achieved through pixelwise

Manuscript received September 17, 2008.

Mohamed El Yazid Boudaren is with the Applied Mathematics Laboratory, EMP, Algiers, Algeria (phone: +213 551 147 274; email: boudaren@gmail.com).

Abdel Belaïd is the head of READ Group, LORIA Research Center, Vandœuvre-lès-Nancy, France (email: abelaid@loria.fr).

naïve Bayesian segmentation. This method considers each pixel taken lonely.

Let us consider a grey level aerial image of a region containing five classes: forest, grass, stone, water and arena which we denote $e_1, e_2 \ldots e_5$ respectively. Thus, for each pixel $s \in S$: y_s and x_s take their values in $V = \{0,1 \ldots 255\}$ and $E = \{e_1 \ldots e_5\}$ respectively.

In the following formulae, we will simply use x instead of x_s and y instead of y_s since we assign a label to each pixel regardless of its location.

The natural object class y of pixel x is chosen so that it maximizes the MAP probability.

$$x^* = \arg\max_{x \in E} P(x/y) = \arg\max_{x \in E} \left[\frac{P(x,y)}{P(y)} \right] \quad (1)$$

Since $P(y)$ is constant while estimating x, we only need to evaluate $P(x,y)$.

$$x^* = \arg\max_{x \in E} P(x,y) \quad (2)$$

For this, we have recourse to Bayes rule.

$$x^* = \arg\max_{x \in E} P(y/x)P(x) \quad (3)$$

The probability distributions $P(x)$ and $P(y/x)$ can be estimated from aerial images samples of the same region. $P(x)$ represents the proportion of the object x in the region under consideration whereas $P(y/x)$ reflects the variability within the object class x.

The criterion adopted here assures a minimal misclassified pixels number when the aerial image is enough big.

However, the resulting mapping when using such a method contains a lot of discontinuities. This is majorly due to noise within the image and the nature of object classes themselves.

Therefore, there are justifiable reasons to think that considering grey level values of neighboring pixels when achieving the classification would improve classification accuracy.

Indeed, when analyzing a pixel taken alone, it is usually impossible even for a human to tell whose class it is.

B. Local contextual Bayesian Segmentation

To surmount the problem discussed above, one can resort to estimating the object class x_s of each pixel $s \in S$ using the local information y_{N_s} where N_s is the neighborhood of pixel s.

$$x_s^* = \arg\max_{x_s \in E} P(x_s / y_{N_s}) = \arg\max_{x_s \in E} \left[\frac{P(x_s, y_{N_s})}{P(y_{N_s})} \right] \quad (4)$$

Nevertheless, the main limit of this method is its heavy computational complexity. In fact, even for a very limited number of object classes N, one can not consider larger than an 8-Neighborhood. Since image size T is usually very big, a huge amount of local information is then overlooked when assigning labels.

Another important drawback of the previous Bayesian methods is that they do not take account of the object class repartition within the aerial image.

C. MRF Bayesian Segmentation

The MRF Bayesian segmentation overcomes the problems discussed in the previous section and enables one to take account of the geographical repartition of natural object classes while segmenting an aerial picture.

The class-map x associated to the aerial image is supposed to be the realization of an MRF.

$$\forall s \in S : P(x_s / x) = P(x_s / x_{N_s}) \quad (5)$$

It is usually computed based on the MAP criterion.

$$x^* = \arg\max_{x \in E^T} P(x/y) \quad (6)$$

As seen before, this is equivalent to:

$$x^* = \arg\max_{x \in E^T} P(y/x)P(x) \quad (7)$$

A particular case of Bayesian labeling consists in using the Hidden Markov Field (HMF) which assumes the following:

$$P(y/x) = \prod_{s \in S} P(y_s / x) \quad (8)$$

$$P(y_s / x) = P(y_s / x_s) \quad (9)$$

Therefore, the likelihood probability becomes:

$$P(y/x) = \prod_{s \in S} P(y_s / x_s) \quad (10)$$

To solve the MAP estimation problem, direct resolutions are intractable. For this reason, many approximation algorithms have been proposed. One of the most popular is ICM algorithm [11].

In [16], Pieczynski suggests to relax the assumption of noise independence by replacing x_s by x_{N_s} in the likelihood probability.

$$P(y/x) = \prod_{s \in S} P(y_s / x_{N_s}) \quad (11)$$

However, the main drawback of HMF based classification in our opinion is that it ties a lonely pixel observation emission y_s to each state x_s whereas one needs usually a larger observation to decide of the class of a pixel.

Indeed, labels can be more accurately associated to regions than to solely pixels. For this reason, several works like [14] consider region-based labeling and start by segmenting the image into regions before assigning labels to pixels.

D. HMM Based Land Cover Mapping

In [1], we proposed a solution based on a hierarchical model constituted of two layers of HMM (figure 1). The first layer comprises one unique HMM we called high-level HMM. It contains as many super states as the number of natural object classes. The high-level HMM models the natural objects geographical repartition. It provides us with the prior probability $P(x)$. Each super state is associated to a low-level HMM that models the corresponding object class. The low-level HMMs constitute the lower layer of the global model that provides us with the likelihood probability $P(y/x)$.

This model principle is very similar to that of NSHP-HMM in the sense that local distributions are tied to high-level model states.

The training of our two layered model has been done in two steps: firstly, the low level HMMs were trained on unitextured pictures. Secondly the high level one was trained on multitextured pictures using the parameters of HMMs of the first step, according to Baum-Welch algorithm.

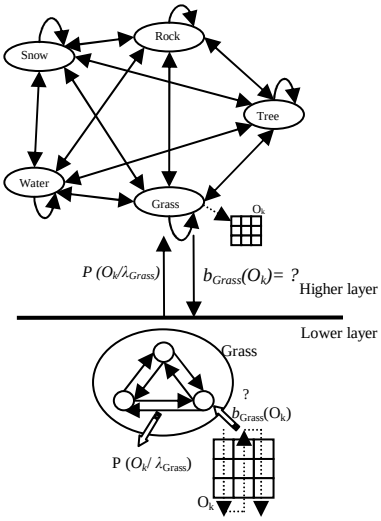

Fig. 1 Two layered HMM architecture

For our experiments, we used real world aerial pictures of a relatively large area, with a resolution of 50 centimeters. Our results were then used to generate reconstituted pictures like depicted in figure 2. This showed that our classifier was able to satisfactorily reproduce the original images.

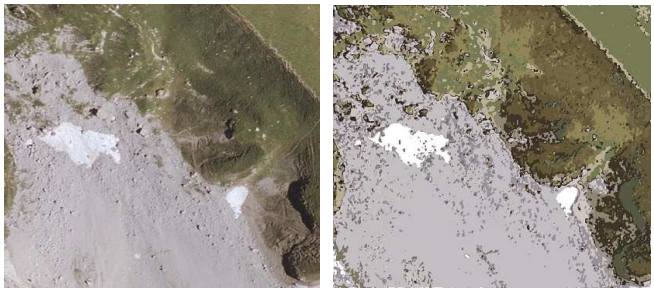

Fig. 2 Aerial image mapping using two layered HMM model: original aerial image (left) and mapping results (right). With the permission of "Régie de Données des Pays de Savoie RGD73-74" [17]

The main weakness of our model is that we scanned aerial images in a linear way. To weaken the repercussion of such a choice, we considered horizontal scanning within high-level HMM and vertical one within low-level models. However, numerous researches proved that 2D modeling gives better results than 1D one.

IV. Combining MRF and HMM for HWR

In [3], we proposed a particularly interesting holistic model called NSHP-HMM which advantageously combines MRF and HMM for handwritten words recognition (HWR). It is mainly based on the use of MRF at pixel level with a switching mechanism between conditional probability distributions assured by an HMM: The HMM analyzes the image columns along the writing axis and an MRF analyzes each column with specific parameters according to the current HMM state. Tying the MRF probability distributions to the HMM states enable the model to dynamically detect local features within the image (strokes of different orientations inherent to handwriting) [3].

In the present section, we briefly present the NSHP-HMM model and show how it is used to model handwritten word images.

A. NSHP-MRF (Non Symmetric Half Plane) Overview

Let L be a lattice of $n \times m$ sites and let $X = \{X_{ij}\}_{(i,j)\in L}$ be a random field defined over the lattice L. In the context of HWR, m and n correspond to the width and length of the word image respectively and each site represents a pixel. X^j stands for the column j of X and $P(X_{ij}/X_A)$, $A \subset L$ stands for $P(X_{ij}/X_{kl}), (k,l) \in A$.

Finally, let us define the sets:

$$\Sigma_{ij} = \{(k,l) \in L / l < j \text{ or } (l = j, k < i)\},\ \Theta_{ij} \subset \Sigma_{ij}$$

Σ_{ij} is called the *non-symmetric half-plane* and Θ_{ij} the support of pixel $(i,j) \in L$.

For binarized images, the symbol set $V = \{0,1\}$. A word image is then a possible realization of a random field.

X is an NSHP-MRF if and only if:

$$P(X_{ij}/X_{\Sigma_{ij}}) = P(X_{ij}/X_{\Theta_{ij}})\ \forall (i,j) \in L \qquad (12)$$

The joint field mass probability $P(X)$ is obtained by:

$$\begin{aligned} P(X) &= \prod_{j=1}^{n} P(X^j/X^{j-1},\ldots X^1) \\ &= \prod_{j=1}^{n}\prod_{i=1}^{m} P(X_{ij}/X_{\Sigma_{ij}}) \\ &= \prod_{j=1}^{n}\prod_{i=1}^{m} P(X_{ij}/X_{\Theta_{ij}}) \end{aligned} \qquad (13)$$

Authors usually the same form of Θ_{ij} for all pixels that is:

$$\Theta = \{\Theta_{ij}\}, (i,j) \in L$$

$$\Theta_{ij} = \{(i-i_k, j-j_k)/1 \le k \le P, j_k > 0 \text{ or } (j_k = 0, i_k > 0)\} \cap L$$

For instance, the pixel support may be equal to:

$$\Theta_{ij} = \{(i-1, j-1), (i, j-1), (i-1, j), (i+1, j-1)\} \cap L$$

B. NSHP-HMM (Non-Symmetric Half-Plane Hidden Markov Model)

The key idea of NSHP-HMM is to tie the conditional probability distributions within image columns to HMM states so that these ones condition the feature sensitivity of the adjacent MRF. For instance, HMM may include specialized states to detect the presence of an upstroke or down-stroke, which would be very advantageous for HWR.

Let us rewrite the joint field mass probability in terms of pattern likelihood with respect to the HMM λ.

$$P(X/\lambda) = \prod_{j=1}^{n} P(X^j/X^{j-1},\ldots X^1,\lambda) = \prod_{j=1}^{n}\prod_{i=1}^{m} P(X_{ij}/X_{\Theta_{ij}},\lambda) \qquad (14)$$

As column X^j is associated to a parallel state stochastic process $Q = q_1 \ldots q_n$.

$$\begin{aligned} P(X/\lambda) &= \sum_Q P(X,Q/\lambda) = \sum_Q P(X/Q,\lambda) P(Q/\lambda) \\ &= \sum_Q \prod_{j=1}^{n} P(q_j/q_{j-1}) P(X^j/X^{j-1}\ldots X^1, q_j, \lambda) \\ &= \sum_Q \prod_{j=1}^{n} P(q_j/q_{j-1}) \prod_{i=1}^{m} P(X_{ij}/X_{\Theta_{ij}}, q_j, \lambda) \end{aligned} \qquad (15)$$

Note that Q is a first order Markov process and that pixel distributions of column X^j depend only on the state q_j.

The results obtain by NSHP-HMM modeling on a real database of unconstrained words are extremely satisfactory. However, in opposition to other MRF based models, NSHP-HMM main weakness is the requiring of image height normalization.

V. AERIAL IMAGES MAPPING USING NSHP-HMF

Our main idea was to adapt the NSHP-HMM modeling to aerial images mapping. However, unlike words images, aerial pictures do not possess a natural ordering of columns. A genuine two-dimensional modeling is then required.

The Hidden Markov Fields (HMF) can be seen as an extension of MRF. They have long been used to model problems where the data is two-dimensional. The basic lack of HMF methods in our opinion is that most of them assume that pixel conditional observations are independent on each other, this is more commonly called the 'noise independence assumption'.

$$P(y/x) = \prod_{s \in S} P(y_s / x_s) \tag{16}$$

where $P(y_s / x_s)$ does not depend on location *s*.

At state x_s level, the observations inter-relationship is then ignored and a huge amount of information is then overlooked. We believe that this kind of information may be of fundamental importance in the context of aerial images mapping and propose to go beyond the noise independence assumption by linking the local observation distribution given by an NSHP-MRF to an HMF states like we did in the NSHP-HMM.

A. Non-Symmetric Half-Plane Hidden Markov Fields

Let L be a lattice of $T = n \times m$ sites and let $X = \{X_{ij}\}_{(i,j) \in L}$ and be $Y = \{Y_{ij}\}_{(i,j) \in L}$ the hidden and the observable random fields respectively of HMF λ defined over the lattice *L*.

In the context of images, sites correspond to pixels. *n* and *m* are image length and width respectively.

If *L* is a grey level image, the symbol set is $V = \{0,1 .. 255\}$.

X_{ij} stands for pixel state and take their values in the state set $E = \{e_1 .. e_N\}$. In the context of aerial images mapping, states correspond to natural object classes.

Y_{ij} stands for pixel (i, j) grey-level value and $P(Y_{ij} / Y_A)$, $A \subset L$ stands for $P(Y_{ij} / Y_{kl}), (k,l) \in A$.

Finally, let us consider the sets:

$$\Sigma_{ij} = \{(k,l) \in L / l < j \text{ or } (l = j, k < i)\}$$
$$\Theta_{ij} \subset \Sigma_{ij}$$
$$\Delta_{ij} \subset \Sigma_{ij}$$

where Σ_{ij} is the *non-symmetric half-plane* and Θ_{ij} and Δ_{ij} are the observable and hidden support respectively of pixel $(i, j) \in L$.

We have:

$$\begin{aligned} P(Y/\lambda) &= \sum_{Y} P(Y, X/\lambda) \\ &= \sum_{Y} P(Y/X, \lambda) P(X/\lambda) \end{aligned} \tag{17}$$

(X, Y) is a Non-Symmetric Half-Plane HMF (NSHP-HMF) if and only if:

- $P(X_{ij} / X, \lambda) = P(X_{ij} / X_{\Delta_{ij}}, \lambda)$. (18)
- $P(Y_{ij} / Y, X, \lambda) = P(Y_{ij} / Y_{\Theta_{ij}}, X_{ij}, \lambda)$. (19)

B. NSHP-2D-HMM Modeling of Aerial Images

To make aerial image modeling through NSHP-HMF possible, we need to define the prior probability $P(X/\lambda)$. For this reason, we will resort in a first time to a special case of HMF models which is 2D-HMM.

2D-HMM is a special case defined in a similar way to 1D-HMM. The only difference is that each state depends on one state in both horizontal and vertical directions. The causality principle is then maintained.

Let us define the state support Δ_{ij}:

$$\Delta_{ij} = \{(i-1, j), (i, j-1)\} \cap L$$

Then, we derive the joint probability:

$$\begin{aligned} P(Y/\lambda) &= \sum_{Y} P(Y, X/\lambda) \\ &= \sum_{Y} P(Y/X, \lambda) P(X/\lambda) \\ &= \sum_{Y} \prod_{(i,j) \in L} P(Y_{ij} / Y_{\Theta_{ij}}, X_{ij}, \lambda) P(X_{ij} / X_{\Delta_{ij}}) \end{aligned} \tag{20}$$

We only need to define Θ_{ij} to make the previous formula fully defined. For instance, we can take:

$$\Theta_{ij} = \{(i-1, j-1), (i, j-1), (i-1, j), (i+1, j-1)\} \cap L$$

More explicitly, the elements of NSHP-2D HMM $\lambda(A, B, \Theta)$ are:

- $\Theta = \{\Theta_{ij}\}, (i, j) \in L$. In this work, we consider: $\Theta_{ij} = \{(i-1, j-1), (i, j-1), (i-1, j), (i+1, j-1)\} \cap L$
- $V = \{v_1 .. v_M\}$, the vocabulary of *M* possible symbols. $\forall (i, j) \in L : Y_{ij} \in V$.

  For instance, we can take $V = \{0 .. 255\}$ for grey-level images.
- $E = \{e_1 ... e_N\}$, the set of *N* possible states of the model. $\forall (i, j) \in L : X_{ij} \in E$.
- $A = \{a_{klm}\}_{1 \le k,l,m \le N}, a_{klm} = P(X_{ij} = e_m / X_{i-1j} = e_k, X_{ij-1} = e_l)$, the state transition matrix.
- $B = \{b_k(y_{ij}, y_{\Theta_{ij}})\}_{1 \le k \le N, (i,j) \in L}$, the conditional pixel observation distribution where $b_k(y_{ij}, y_{\Theta_{ij}}) = P(Y_{ij} = y_{ij} / Y_{\Theta_{ij}} = y_{\Theta_{ij}}, X_j = e_k)$.

Let us now consider the aerial image mapping problem in the NSHP-2D-HMM context: given the observation Y of an aerial image *L* and a NSHP-2D-HMM $\lambda(A, B, \Theta)$, mapping consists in assigning the corresponding X^* that maximizes the MAP probability.

$$\begin{aligned} X^* &= \arg\max_{x \in E^T} P(Y/X) P(X) \\ &= \arg\max_{x \in E^T} \prod_{(i,j) \in L} P(Y_{ij} / Y_{\Theta_{ij}}, X_{ij}, \lambda) P(X_{ij} / X_{\Delta_{ij}}) \\ &= \arg\max_{x \in E^T} \prod_{(i,j) \in L} b_{X_{ij}}(Y_{ij}, Y_{\Theta_{ij}}) a_{X_{\Delta_{ij}} X_{ij}} \end{aligned} \tag{21}$$

Where the state set $E = \{e_1 ... e_N\}$ is the set of natural object classes and the symbol set $V = \{0 .. 255\}$ if we consider grey-level aerial images.

Finding out X^* is an NP-hard problem [15]. In fact, even for classical 2D-HMM, the optimal decoding procedure is

intractable in practice; contrary to 1D-HMM, the factorization of computation is not possible in 2D-HMM.

Fortunately, relaxation methods that yield good results exist in literature [4, 5]. In [5], an interesting new model called Dependency Tree Hidden Markov Model (DT-HMM) was proposed to overcome the complexity problem of 2D-HMM while keeping the two-dimensional aspect of data. It is mainly based on the idea that each state depends on only one neighboring state at a time.

Thereafter, we summarize the main aspects of DT-HMM. Then, we will develop the DT-NSHP-HMM. Finally, we will derive the training and recognition procedures and show that they exhibit a reasonable computational complexity.

C. DT-HMM Overview

The main idea of DT-HMM is that each state depends on only one neighboring state at the time. This neighboring state may be the horizontal or the vertical one depending on a direction random variable $t(i,j)$ such that:

$$t(i,j)=\begin{cases}(i-1,j)\ with\ \ prob\ \ 0.5\\(i,j-1)\ with\ \ prob\ \ 0.5\end{cases} \tag{22}$$

The model assumes the following:

$$P\left(X_{ij}/X_{\Delta_{ij}}\right)=\begin{cases}P_V\left(X_{ij}/X_{i-1\,j}\right) if\ \ t(i,j)=(i-1,j)\\P_H\left(X_{ij}/X_{i-1\,j}\right) if\ \ t(i,j)=(i,j-1)\end{cases} \tag{23}$$

where $\Delta_{ij}=\{(i-1,j),(i,j-1)\}\cap L$.

Let the direction function be:

$$D(i,j)=\begin{cases}V & if\ \ t(i,j)=(i-1,j)\\H & if\ \ t(i,j)=(i,j-1)\end{cases} \tag{24}$$

Consequently,

$$P\left(X_{ij}/X_{\Delta_{ij}}\right)=P_{D(i,j)}\left(X_{ij}/X_{t(i,j)}\right) \tag{25}$$

The variable *t* defines a tree structure (dependency tree) over the lattice *L* with pixel (0,0) as the root.

D. DT-NSHP-HMM Definition

The DT-NSHP-HMM has the same parameters as the NSHP-2D-HMM. The only difference is in *A* matrix.

Horizontal and vertical states transitions are estimated separately. *A* matrix is then replaced by two matrices that we denote A_H and A_V where:

$$\begin{aligned}A_H&=\{a_{kl}\}_{1\le k,l\le N}, a_{kl}=P\left(X_{ij}=e_l/X_{ij-1}=e_k\right)\\A_V&=\{a_{kl}\}_{1\le k,l\le N}, a_{kl}=P\left(X_{ij}=e_l/X_{i-1j}=e_k\right)\end{aligned} \tag{26}$$

For the sake of simplicity, we will denote the DT-NSHP-HMM $\lambda(A,B,\Theta)$ where $A=\{A_H,A_V\}$.

The image likelihood calculus given a dependency tree is as follows:

$$\begin{aligned}P(Y/\lambda,t)&=\sum_X P(Y,X/\lambda)\\&=\sum_X P(Y/X,\lambda,t)P(X/\lambda,t)\\&=\sum_X\prod_{(i,j)\in L} b_{X_{ij}}\left(Y_{ij},Y_{\Theta_{ij}}\right)a_{D(i,j)}\left(X_{t(i,j)},X_{ij}\right)\end{aligned} \tag{27}$$

E. Labeling Procedure

Given the observation y of an aerial image *L* and a DT-NSHP-HMM $\lambda(A,B,\Theta)$, mapping consists in assigning the corresponding *x* that maximizes the MAP probability given a dependency tree function *t*.

$$x^*=\arg\max_{x\in E^T} P(y/x)P(x) \tag{28}$$

In the following we will adapt the DT-HMM *Viterbi* procedure to the DT-NSHP-HMM model.

In a first time, we assume the dependency tree given.

Let $T(i,j)$ be the sub-tree having pixel (i,j) as a root and let $\beta_{i,j}(k)$ be the maximum probability that $Y_{T(i,j)}$ is generated starting by state *k* in the root (i,j).

Let us define:

$$H_k(i,j)=\begin{cases}\max\limits_{l\in E} a_H(k,l)\beta_{ij+1}(l) if\ \ t(i,j+1)=(i,j)\\1 \qquad\qquad otherwise\end{cases} \tag{29}$$

$$V_k(i,j)=\begin{cases}\max\limits_{l\in E} a_V(k,l)\beta_{i+1j}(l) if\ \ t(i+1,j)=(i,j)\\1 \qquad\qquad otherwise\end{cases} \tag{30}$$

Thereafter, let us consider:

$$D_k(i,j)=H_k(i,j)V_k(i,j) \tag{31}$$

We can compute the values of $\beta_k(i,j)$ recursively as follows:

$$\beta_k(i,j)=b_k\left(y_{ij}/y_{\Theta_{ij}},\phi_k\right)D_k(i,j) \tag{32}$$

Finally, let us define:

$$x^*_{T(i,j)}=\arg\max_{X_{T(i,j)}}\beta_{i,j}(k) \tag{33}$$

Note that $x^*_{T(0,0)}=\arg\max_{X_{T(0,0)}}\beta_{0,0}(k)$ is the solution of the MAP labeling defined above.

$$x^*=x^*_{T(0,0)} \tag{34}$$

Thereafter, we show how the previous algorithm can be used iteratively to produce the image labeling.

According to the MAP criterion, we have:

$$\begin{aligned}X^*&=\arg\max_{X\in E^T} P(Y,X)\\&\approx\arg\max_{X\in E^T}\sum_t P(Y,X/t)\end{aligned} \tag{35}$$

Let us assume:

$$\sum_t P(Y,X/t)\approx\max_t P(Y,X/t) \tag{36}$$

Consequently,

$$\begin{aligned}X^*&\approx\arg\max_{X\in E^T}\left\{\max_t P(Y,X/t)\right\}\\&\approx\arg\max_{X\in E^T}\left\{\max_t P(Y/X,t)P(X/t)\right\}\end{aligned} \tag{37}$$

As $P(Y/X,t)=P(Y/X)$, therefore:

$$X^*\approx\arg\max_{X\in E^T}\left\{P(Y/X)\max_t P(X/t)\right\} \tag{38}$$

Let us denote $t^*=\arg\max_t P(X/t)$ (39)

We get:

$$X^*\approx\arg\max_{X\in E^T} P(Y/X,t)P(X/t^*) \tag{40}$$

Which we propose to solve iteratively by maximizing over *t* and *X* alternatively like follows:

$$\begin{cases}X^*=\arg\max\limits_{X\in E^T} P(Y/X)P(X/t^*)\\t^*=\arg\max\limits_t P(X^*/t)\end{cases} \tag{41}$$

Hence, an initialization is required to run the iterative process. We can choose either to start by initializing the dependency tree or the labeling.

- Initialization: Initialize Dependency tree:

$$t(i,j)=\arg\min_{t(i,j)}\left\|y_{(i,j)}-y_{t(i,j)}\right\|$$

where $\| \ \|$ is Euclidian distance.

- Step 1: Achieve adapted Viterbi alignment as described above.
- Step 2: Update Dependency tree as follows:
$$t(i,j) = \arg\max_{t(i,j)} a_{D(i,j)}(x_{t(i,j)}, x_{(i,j)})$$
- Step 3: If end criterion not reached go to step 1.
- End

F. Training

To train our model on aerial images, we suggest performing the learning in two steps like we did in [1].

First, to compute the observations conditional probabilities, we devote an NSHP-MRF model to each natural object class. We train each model on a natural object class taken alone. Explicitly, we need to have unitextured aerial images (a set of aerial images per object class). Then, we estimate the state transition matrix *A* using the parameters of the previous learned NSHP-MRF models.

Let ϕ_k be the NSHP-MRF corresponding to object class of state e_k. Therefore, the classical emission probability given by $b_k(y_{ij}, y_{\Theta_{ij}})$ is computed as follows:

$$b_k(y_{ij}, y_{\Theta_{ij}}) = P(y_{ij} / y_{\Theta_{ij}}, \phi_k) \quad (42)$$

To estimate the transition matrix, we assume in a first time that we have labeled aerial pictures.

Thus, we can simply resort to a frequency-based training.

Let us define the parameters:

$$\begin{aligned} \zeta(k)_{1\le k\le N} &= \sum_{(i,j)\in L} 1_{(x_{ij}=e_k)} \\ \zeta_V(k,l)_{1\le k,l\le N} &= \sum_{(i,j)\in L} 1_{(x_{ij}=e_l, x_{t(i-1,j)}=e_k)} \\ \zeta_H(k,l)_{1\le k,l\le N} &= \sum_{(i,j)\in L} 1_{(x_{ij}=e_l, x_{t(i,j-1)}=e_k)} \end{aligned} \quad (43)$$

We derive then the matrix *A* parameters as follows:

$$a_H(k,l) = \frac{\zeta_H(k,l)}{\zeta(k)},\ a_V(k,l) = \frac{\zeta_V(k,l)}{\zeta(k)} \quad (44)$$

The computational complexity order of this method is of *T* since we can compute the previous parameters while analyzing the image pixel by pixel, whereas the memory complexity is of $2N^2$. *T* must be enough big to accurately model the real relationships between natural objects.

If labeled images do not exist, we can estimate the state transition matrix in an iterative manner as follows:

- Initialization: Initialize transition matrix A uniformly.
- Step 1: Achieve Viterbi alignment. This provides an image labeling.
- Step 2: Re-estimate *A* matrix as described above.
- Step 3: If end criterion not reached go to step 1.
- End

G. Model Complexity

In this section, we demonstrate that our model exhibits a reasonable computational complexity.

Let us consider a DT-NSHP-HMM $\lambda(A, B, \Theta)$ and an aerial image *L* of size *T*. Let *N* be the number of states and *M* the number of symbols.

NSHP-MRF training is performed independently on unitextured images. Since, this kind of modeling already exist, there is no need to analyze its complexity.

DT-HMM modeling also already exists and it was shown in [6] that its complexity is linear with the image size *T*.

Accordingly, we only need to analyze our adapted Viterbi procedure. We will prove that we can convert our model to a simple DT-HMM after a set of reasonable-cost computations.

If we compute for each pixel $(i, j) \in L$ and each state $k \in E$ the value of $b_k(y_{ij}, y_{\Theta_{ij}})$, we will have all the *B* matrix elements necessary to apply Viterbi procedure in a simple DT-HMM context.

Consequently, the model complexity is tractable in practice.

VI. Conclusion

In this paper, we addressed the problem of land cover mapping in aerial images.

After reviewing Markov models based previous works, we reminded the NSHP-HMM formalism that we proposed in [3] for HWR and defined three new NSHP-like models.

First, we defined the NSHP-HMF which is a special case of HMF that assumes observations local conditional dependence within aerial image.

Then, we defined a particular NSHP-HMF that we called NSHP-2D-HMM which can be seen as an extension of the NSHP-HMM model since it assign a state to each image pixel instead of a whole image column.

Like HMF and 2D-HMM, direct inference methods of our new models are intractable in practice. However, approximation and simulation techniques may be used. In this paper, we proposed to approximate the NSHP-2D-HMM by a DT-NSHP-HMM that extends the DT-HMM.

Viterbi algorithm was defined in a similar way to DT-HMM. The only difference is in *B* matrix elements computation. We also defined a new labeling algorithm based on an iterative dependency tree construction.

Note that, contrary to NSHP-HMM and NSHP-HMF which can be considered as special cases of HMM and HMF respectively, NSHP-2D-HMM and DT-NSHP-HMM are extensions of 2D-HMM and DT-HMM respectively.

Contrary to handwritten recognition problem, there is no need for image normalization in image labeling here given the nature of aerial images.

As future work, we propose to apply DT-NSHP-HMM to handwritten problem. Further study can be devoted to dependency tree training given the nature of words: same words have more likely a same dependency tree.

References

[1] Mohamed El Yazid Boudaren, Abdenour Labed, Adel Aziz Boulfekhar and Yacine Amara, *Hidden Markov model based classification of natural objects in aerial pictures*, Proceeding of IAENG International Conference on Signal and Image Engineering, London, July 2-4, 2008.

[2] Mohamed El Yazid Boudaren, Abdenour Labed, Adel Aziz Boulfekhar and Yacine Amara, *HMM based classification of natural objects in aerial pictures*, Abstract Volume of GFKL, Hamburg, Germany, July 16-18, 2008.

[3] G. Saon, A, Belaïd, *High performance unconstrained word recognition system combining HMMs and Markov random fields*, Volume 28, World Scientific, 1997.

[4] B. Merialdo, S. Marchand-Maillet, B. Huet, *Approximate Viterbi decoding for 2D-hidden Markov models*, IEEE International

Conference on Acoustics, Speech and Signal Processing, Volume 6, 5-9 June 2000 , pp. 2147 – 2150.

[5] B. Merialdo, *Dependency Tree Hidden Markov Models*, Research Report RR-05-128, Institut Eurecom, 2005.

[6] B. Merialdo, J. Jiten, E. Galmar, B. Huet, *A new approach to probabilistic image modeling with multidimensional hidden Markov models*, Adaptive multimedia retrieval 2006, pp. 95-107.

[7] Xavier Descombes, Miguel Moctezuma, Henri Maître and Jean-Paul Rudant, *Coastline detection by a Markovian segmentation on SAR images*, Signal Processing, Volume 55, 1996, pp. 123-132.

[8] Teerasit Kasetkasem, Manoj K. Arora and Pramod K. Varshney, *Super-resolution land cover mapping using a Markov random field based approach*, Remote Sensing of Environment, Volume 96, Elsevier, 2005, pp. 302-314.

[9] Chi Hau Chen, Pei-Gee Peter Ho, *Statistical pattern recognition in remote sensing*, Pattern Recognition, Volume 41, Elsevier, 2008, pp. 2731-2741.

[10] Desheng Liu, Kuan Song, John R. G. Townshend and Peng Gong, *Using local transition probability models in Markov random fields for forest change detection*, Remote Sensing of Environment, Volume 112, Elsevier, 2008, pp. 2222-2231.

[11] J.E. Besag, *On the statistical analysis of dirty pictures (with discussion)*, J. Roy. Statist. Soc. (B) 48, 1986, pp. 259–302.

[12] L. Zhang, X. Wu, *An edge guided image interpolation algorithm via directional filtering and data fusion*, IEEE Transactions on Image Processing 15 (8), 2006, pp. 2226–2238.

[13] C.M. Lai, K.M. Lam, W.C. Siu, *An efficient fractal-based algorithm for image magnification*, Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing, October 2004, pp. 571–574.

[14] Roger Trias Sanz, Didier Boldo, *A high-reliability, high-resolution method for land cover classification into forest and non-forest*, 14th Conference on Image Analysis, Finland, 2005.

[15] E. Levin, R. Pieraccini, *Dynamic planar warping for optical character recognition*, IEEE International Conference on Acoustics, Speech and Signal Processing, Volume 3, 23-26 March 1992, pp.149 – 152.

[16] Wojciech Pieczynski, *Markov models in image processing*, Traitement de Signal, Volume 20 N°3, 2003, pp.255-277.

[17] RGD73-74: Régie de Gestion des Données des Deux Savoies. http://www.rgd73-74.fr.